Evaluating the Quality, Usability, and Fitness of Open Data for Public Health Research

Please Dial Conference Phone: 877-394-0659; Meeting Code: 775 483 8037#.
Please mute your phone and computer speakers during the presentation.
You may download today’s presentation and speaker bios from the ‘Files 2’ box at the top right corner of your screen.
Agenda

Welcome: Rick Ingram, DrPH, PHSSR National Coordinating Center, Assistant Professor, U. of Kentucky College of Public Health

Presenter:
“Evaluating the Quality, Usability, and Fitness of Open Data for Public Health Research”
Erika G. Martin, PhD, MPH, Assistant Professor, Public Administration and Policy, Rockefeller College of Public Affairs and Policy, SUNY – Albany

Commentary:
Guthrie Birkhead, MD, MPH, Deputy Commissioner, Office of Public Health, New York State Department of Health
Cheryl Wold, MPH, Wold and Associates, Pasadena, California

Questions and Discussion

Future Webinar Announcements
PHSSR Mentored Researcher Development Awards

- 2-year awards providing protected time to complete PHSSR project, with research mentor and practice mentor (2013-2015)
- Four award recipients presenting in the series

<table>
<thead>
<tr>
<th>Topic</th>
<th>Speaker</th>
<th>Institution</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Identifying &amp; Learning from Positive Deviant Local Public Health Departments in Maternal and Child Health</td>
<td>Tamar A. Klaiman, PhD, MPH, U. of Sciences, Philadelphia</td>
<td>(February 19)</td>
<td></td>
</tr>
<tr>
<td>Leveraging Electronic Health Records for Public Health: <em>From Automated Disease Reporting to Developing Population Health Indicators</em></td>
<td>Brian Dixon, PhD, Indiana University</td>
<td>(March 4)</td>
<td></td>
</tr>
<tr>
<td>Evaluating the Quality, Usability, and Fitness of Open Data for Public Health Research</td>
<td>Erika G. Martin, PhD, MPH, State University of New York - Albany</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Restructuring a State Nutrition Education and Obesity Prevention Program: <em>Implications of a Local Health Department Model</em></td>
<td>Helen W. Wu, PhD, U. California - Davis</td>
<td>(April 1)</td>
<td></td>
</tr>
</tbody>
</table>
Presenter

Erika G. Martin, PhD, MPH

Assistant Professor, Public Administration and Policy, Rockefeller College of Public Affairs and Policy

Senior Fellow and Director of Health Policy Studies, Nelson A. Rockefeller Institute of Government

University at Albany, State University of New York

2013 PHSSR Mentored Researcher Development Award Recipient

erika.gale.martin@gmail.com
Evaluating the Quality, Usability, and Fitness of Open Health Data for Public Health Research

Erika Martin, PhD MPH
Rockefeller Institute of Government & University at Albany

PHSSR Research-in-Progress Webinar
March 11, 2015
Acknowledgements & Disclosures

- Funding from the Robert Wood Johnson Foundation’s Public Health Services & Systems Research Program (grant ID #71597 to Martin and Birkhead)
- Coauthors: Gus Birkhead, Natalie Helbig, Jennie Law, Weijia Ran
- Early feedback: Courtney Burke, Patricia Lynch, Theresa Pardo, Ozlem Uzuner
- JSON technical support: Chris Kotfila
- Gus Birkhead and Natalie Helbig are employees of the New York State Department of Health, which maintains the Health Data NY open data platform reviewed in this study
Promises of open data

Research and practice gaps
- Making open data usable and high quality for public health research
- Research methods to document characteristics of open data offerings and differences across platforms
  - Sampling design
  - Coding instrument
  - Statistical analysis

Findings and implications for practice

Future project activities
Open data background

- New source of information for public health research
- Motivated by government transparency movement, including President Obama’s memorandum on open government
- Thousands of government datasets released on open data platforms at federal, state, and local levels meeting several “openness” criteria
  - Publicly accessible, available in non-proprietary formats, free of charge, unlimited use and distribution rights
- New opportunities for public health research and practice
  - New York State examples in Martin, Helbig, Shah *JAMA* 2014
Search engines to locate data objects
Capabilities to interact directly with data in the platform
Challenges and resources for developers

New York State Health Data Code@thon
Where Innovation Meets Public Health

NYS Health Innovation Challenge

Submission Deadline: July 31, 2014

Prizes
- First Place $30,000
- Second Place $10,000
- Third Place $3,000

Pre-Register

Build something awesome with Open Data!

The Socrata Open Data API allows you to programatically access a wealth of open data resources from governments, non-profits, and NGOs around the world. Click the link below and try a live example right now.


App Developers
Looking to use open data as part of your application or your business? Learn how to get started.

Libraries & SDKs
Support for most popular programming languages and platforms.

Need Help?
Struggling with a problem you can't figure out? Get help fast!
Opportunities to submit ideas for new datasets and provide user feedback
Research questions

- Open data are promising but...

- To what extent are open health data **usable** and **fit** for public health research?

- How could government agencies improve the **quality** of the **data** and corresponding **metadata**, to make these data more usable and fit for public health researchers and practitioners?
Research design overview

- Systematic review of open health data offerings on federal, state, and local platforms
  - Adapted from Institute of Medicine and Patient-Centered Outcomes Research Institute guidelines for systematic literature reviews
- Health-related data offerings randomly sampled from three platforms
  - Healthdata.gov (federal)
  - Health Data NY (state)
  - NYC Open Data (city)
- All data offerings examined with a coding guide to evaluate:
  - Data quality (intrinsic, contextual)
  - Five-star open data deployment
  - Metadata quality
  - Platform usability
Sampling design

- **Final selection**
  - All NYC Open Data offerings related to health (N=37)
  - 25% random sample of Health Data NY data objects (N=71)
  - 5% random sample of Healthdata.gov data objects (N=75)
  - Total of 183 data objects

- **Systematic random sampling of data offerings**
  - Metadata from platforms scraped into three Excel spreadsheets
  - Excel-based random number generator assigned random integer values from 1 to N, then selected every dataset assigned a 1
Development of coding guide

- Cross-disciplinary literature review to develop a preliminary conceptual framework of data quality, usability, and fitness

- Stakeholder conversations to refine conceptual framework
  - Respondents: experts in computer science/semantic web (1) and data quality (2); academic health researchers (3); local health department epidemiologists (3); analysts at health policy and advocacy center (2)
  - Topics covered: how health data are used; which health datasets are useful; how respondents decide whether a dataset is of high quality, usable, and fit; metadata needed to evaluate datasets; comments on conceptual framework

- Internal vetting with interdisciplinary research team
Development of coding guide, cont.

- Additional stakeholder input on the quality, usability, and fitness of data for health research obtained from:
  - Focus groups of public health researchers and practitioners, conducted at November 2013 open data workshop in Albany, NY (Martin, Helbig, Birkhead *J Public Health Manag Pract* 2014)
  - Blog post to NYSDOH SAS user group to solicit comments
  - Review of stakeholder feedback comments on the Prevention Agenda dashboard
  - Review of a sample of data-based County Health Assessments
  - Grant reviewers’ feedback

- Extensive pilot-testing and refinement
Categories of questions

- Descriptive information
- Intrinsic data quality
- Contextual data quality
- Adherence to Dublin Core international metadata standards
- Consistency with five-star open data deployment scheme
Dublin Core international metadata standards

<table>
<thead>
<tr>
<th>Term Name:</th>
<th>contributor</th>
</tr>
</thead>
<tbody>
<tr>
<td>URI:</td>
<td><a href="http://purl.org/dc/elements/1.1/contributor">http://purl.org/dc/elements/1.1/contributor</a></td>
</tr>
<tr>
<td>Label:</td>
<td>Contributor</td>
</tr>
<tr>
<td>Definition:</td>
<td>An entity responsible for making contributions to the resource.</td>
</tr>
<tr>
<td>Comment:</td>
<td>Examples of a Contributor include a person, an organization, or a service. Typically, the name of a Contributor should be used to indicate the entity.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Term Name:</th>
<th>coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>URI:</td>
<td><a href="http://purl.org/dc/elements/1.1/coverage">http://purl.org/dc/elements/1.1/coverage</a></td>
</tr>
<tr>
<td>Label:</td>
<td>Coverage</td>
</tr>
<tr>
<td>Definition:</td>
<td>The spatial or temporal topic of the resource, the spatial applicability of the resource, or the jurisdiction under which the resource is relevant.</td>
</tr>
<tr>
<td>Comment:</td>
<td>Spatial topic and spatial applicability may be a named place or a location specified by its geographic coordinates. Temporal topic may be a named period, date, or date range. A jurisdiction may be a named administrative entity or a geographic place to which the resource applies. Recommended best practice is to use a controlled vocabulary such as the Thesaurus of Geographic Names [TGN]. Where appropriate, named places or time periods can be used in preference to numeric identifiers such as sets of coordinates or date ranges.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Term Name:</th>
<th>creator</th>
</tr>
</thead>
<tbody>
<tr>
<td>URI:</td>
<td><a href="http://purl.org/dc/elements/1.1/creator">http://purl.org/dc/elements/1.1/creator</a></td>
</tr>
<tr>
<td>Label:</td>
<td>Creator</td>
</tr>
<tr>
<td>Definition:</td>
<td>An entity primarily responsible for making the resource.</td>
</tr>
<tr>
<td>Comment:</td>
<td>Examples of a Creator include a person, an organization, or a service. Typically, the name of a Creator should be used to indicate the entity.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Term Name:</th>
<th>date</th>
</tr>
</thead>
<tbody>
<tr>
<td>URI:</td>
<td><a href="http://purl.org/dc/elements/1.1/date">http://purl.org/dc/elements/1.1/date</a></td>
</tr>
</tbody>
</table>

http://dublincore.org/documents/dces/
Five-star open data deployment scheme

 ★ make your stuff available on the Web (whatever format) under an open license

 ★ ★ make it available as structured data (e.g., Excel instead of image scan of a table)

 ★ ★ ★ use non-proprietary formats (e.g., CSV instead of Excel)

 ★ ★ ★ ★ use URIs to denote things, so that people can point at your stuff

 ★ ★ ★ ★ ★ link your data to other data to provide context

http://5stardata.info/

OL = OnLine
RE = can be REused
OF = Open Formats
URI: Uniform Resource Identifier
LD = can Link Data
Example of coding guide questions

- Contextual data quality – ease of manipulation
  - What is the data object’s primary presentation format (table, chart, map, external file, application programming interface (API), filter, other)?
  - If primary format is a visualization, are simple statistics available?
  - Are there different presentation formats for the data object (if so, list available formats)?
  - Can the data be downloaded from the platform (if so, what download options are available)?
  - Can the data be downloaded from the data access page (if so, what download options are available)?
  - Are the data available as structured data?
  - Are the data available in non-proprietary formats?
  - Is the selection a data artifact?
  - Is the data object viewable in a browser (if no, why not)?
Intrinsic data quality – accuracy/objectivity/reliability

- Is a limitations section clearly and explicitly identified?*
- Is there a codebook or data dictionary?
- Is any information about the purpose of the data collection listed?*
- Is there a description of the sample design?*
- Is there a description of how the data were collected?*
- Is the data collection instrument available?*
- Is there any notation about random checks for data accuracy, auditing procedures, validity checks, etc.?*
- Is there any notation about the data preparation/processing steps that happened as the data were transformed into open data?*

* if yes, coders copy and paste relevant text
Example of coding guide questions, cont.

- Contextual data quality – relevancy/value-added
  - Is there a data object description?*
  - Is the granularity clearly and specifically identified?*
  - Is the unit of analysis clearly and specifically identified?*
  - Is the data object available via a uniform resource identifier (URI) on the metadata page?*
  - Are there examples of how data have been used in research/practice?*
  - Does the platform list any ideas for how data could be used?*
  - Is there mention of other data objects that would be of interest?*
  - Are the data available in resource descriptive framework (RDF) format?*
  - Do variable names hyperlink to contextual information?*
  - Series of questions on presence of demographic, provider, and health facility variables, and their response categories
    - Demographics: age, gender, race/ethnicity, insurance status, income, education

* if yes, coders copy and paste relevant text
Additional coding guide considerations

- Static documents archived on hard drive
  - Codebooks, data dictionaries, dataset downloads, other available materials online
  - Metadata and data access pages saved as complete webpages

- Questions very specific and direct, to improve inter-rater reliability
Data collection procedures

- Extensive pilot-testing of coding guide
  - Purposive selection of 16 data offerings from the three platforms which varied widely (e.g. administrative data vs survey, simple tabular format vs large SAS-file download, small vs large size)
  - J.L. and W.R. double-coded and compared responses, discussing discrepancies with E.M.
  - Interim feedback from N.H. and G.B.
  - Coding guide continuously updated until uniform agreement

- Coding guide transformed into Access database for data entry
  - Form view and fixed response categories to minimize data entry errors
  - Flags for queries to discuss with the team

- Separate coding guide for platform usability
  - Assessed after all offerings coded
Main findings

- Only one-quarter of open data offerings are tabular datasets

- Most offerings do not contain demographic variables commonly used in public health research

- Health Data NY scored highest on intrinsic data quality, contextual data quality, and adherence to Dublin Core metadata standards

- Gaps in meeting “open data” deployment criteria
  - All offerings met basic “web availability” open data standards
  - Fewer met higher standards of being hyperlinked to other data
<table>
<thead>
<tr>
<th>Characteristic</th>
<th>NYC Open Data (city, N=38)</th>
<th>Health Data NY (state, N=71)</th>
<th>Healthdata.gov (federal, N=74)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Primary presentation format in web browser, N (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Table</td>
<td>17 (44.7)</td>
<td>17 (23.9)</td>
<td>12 (16.2)</td>
</tr>
<tr>
<td>Chart</td>
<td>--</td>
<td>27 (38.0)</td>
<td>--</td>
</tr>
<tr>
<td>Map</td>
<td>9 (23.7)</td>
<td>10 (14.1)</td>
<td>1 (1.4)</td>
</tr>
<tr>
<td>External file</td>
<td>1 (2.6)</td>
<td>9 (12.7)</td>
<td>27 (36.5)</td>
</tr>
<tr>
<td>Application programming interface</td>
<td>--</td>
<td>2 (2.8)</td>
<td>1 (1.4)</td>
</tr>
<tr>
<td>Query tool</td>
<td>4 (10.5)</td>
<td>2 (2.8)</td>
<td>8 (10.8)</td>
</tr>
<tr>
<td>Documents about data</td>
<td>3 (7.9)</td>
<td>1 (1.4)</td>
<td>18 (24.3)</td>
</tr>
<tr>
<td>Not viewable in a browser</td>
<td>4 (10.5)</td>
<td>3 (4.2)</td>
<td>7 (9.5)</td>
</tr>
<tr>
<td><strong>Availability of additional presentation formats, N (%)</strong></td>
<td>11 (29.0)</td>
<td>42 (59.2)</td>
<td>10 (13.5)</td>
</tr>
<tr>
<td><strong>Availability of data related to visualizations, N (%)</strong></td>
<td>5 (55.6)</td>
<td>34 (91.9)</td>
<td>1 (100.0)</td>
</tr>
<tr>
<td><strong>Ability to view data object in browser, N (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Object is viewable in a browser</td>
<td>28 (73.7)</td>
<td>56 (78.9)</td>
<td>27 (36.5)</td>
</tr>
<tr>
<td>Problem with the data access page</td>
<td>5 (13.2)</td>
<td>1 (1.4)</td>
<td>5 (6.8)</td>
</tr>
<tr>
<td>Data object is an external file</td>
<td>2 (5.3)</td>
<td>13 (18.3)</td>
<td>21 (28.4)</td>
</tr>
<tr>
<td>Data object requires subscription or registration</td>
<td>1 (2.6)</td>
<td>--</td>
<td>6 (8.1)</td>
</tr>
<tr>
<td>Data object is only viewable in a proprietary format</td>
<td>1 (2.6)</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>Data object not downloadable for other reasons</td>
<td>1 (2.6)</td>
<td>1 (1.4)</td>
<td>15 (20.3)</td>
</tr>
<tr>
<td><strong>Ability to download data, N (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Available via platform</td>
<td>--</td>
<td>10 (14.1)</td>
<td>--</td>
</tr>
<tr>
<td>Available via data access page</td>
<td>--</td>
<td>--</td>
<td>19 (25.7)</td>
</tr>
<tr>
<td>Available from both sources</td>
<td>32 (84.2)</td>
<td>56 (78.9)</td>
<td>23 (31.1)</td>
</tr>
<tr>
<td>Not available for download</td>
<td>6 (15.8)</td>
<td>5 (7.0)</td>
<td>32 (43.2)</td>
</tr>
<tr>
<td>Characteristic</td>
<td>NYC Open Data (city, N=38)</td>
<td>Health Data NY (state, N=71)</td>
<td>Healthdata.gov (federal, N=74)</td>
</tr>
<tr>
<td>----------------------------------------------------</td>
<td>-----------------------------</td>
<td>-----------------------------</td>
<td>-------------------------------</td>
</tr>
<tr>
<td><strong>Data object year</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Historical data, (^4) (N (%))</td>
<td>12 (31.6)</td>
<td>31 (43.7)</td>
<td>22 (29.7)</td>
</tr>
<tr>
<td>Includes multiple years, (N (%))</td>
<td>7 (18.4)</td>
<td>38 (53.5)</td>
<td>13 (17.6)</td>
</tr>
<tr>
<td><strong>Data update frequency, (N (%))</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Daily or Weekly</td>
<td>1 (2.6)</td>
<td>3 (4.2)</td>
<td>--</td>
</tr>
<tr>
<td>Monthly</td>
<td>3 (7.9)</td>
<td>8 (11.3)</td>
<td>1 (5.3)</td>
</tr>
<tr>
<td>Quarterly, semi-quarterly, or biannually</td>
<td>2 (5.3)</td>
<td>7 (9.9)</td>
<td>5 (26.3)</td>
</tr>
<tr>
<td>Annually or biennially</td>
<td>3 (7.9)</td>
<td>50 (70.4)</td>
<td>8 (42.1)</td>
</tr>
<tr>
<td>As needed</td>
<td>20 (52.6)</td>
<td>1 (1.4)</td>
<td>--</td>
</tr>
<tr>
<td>Not reported</td>
<td>3 (7.9)</td>
<td>1 (1.4)</td>
<td>59 (79.7)</td>
</tr>
<tr>
<td>Not updated</td>
<td>6 (15.8)</td>
<td>1 (1.4)</td>
<td>1 (1.4)</td>
</tr>
<tr>
<td><strong>Inclusion of demographic variables, (N (%))</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>2 (5.3)</td>
<td>21 (29.6)</td>
<td>18 (24.3)</td>
</tr>
<tr>
<td>Gender</td>
<td>2 (5.3)</td>
<td>13 (18.3)</td>
<td>14 (18.9)</td>
</tr>
<tr>
<td>Race/ethnicity</td>
<td>2 (5.3)</td>
<td>8 (11.3)</td>
<td>10 (13.5)</td>
</tr>
<tr>
<td>Insurance status</td>
<td>2 (5.3)</td>
<td>20 (28.1)</td>
<td>18 (24.3)</td>
</tr>
<tr>
<td>Education</td>
<td>2 (5.3)</td>
<td>10 (14.0)</td>
<td>2 (2.7)</td>
</tr>
<tr>
<td>Income</td>
<td>7 (18.4)</td>
<td>5 (7.0)</td>
<td>8 (10.8)</td>
</tr>
<tr>
<td>Geographic identifier</td>
<td>17 (44.7)</td>
<td>45 (63.4)</td>
<td>28 (37.8)</td>
</tr>
<tr>
<td>Provider and/or health facilities</td>
<td>18 (47.4)</td>
<td>36 (50.7)</td>
<td>24 (32.4)</td>
</tr>
<tr>
<td><strong>Size of data object, (^3) median (IQR)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of rows</td>
<td>11 (69)</td>
<td>161 (3340)</td>
<td>357 (2011)</td>
</tr>
<tr>
<td>Number of columns</td>
<td>6 (4)</td>
<td>18 (8)</td>
<td>11 (17)</td>
</tr>
<tr>
<td><strong>Data object hosted on a different platform, (^6) (% (N))</strong></td>
<td>n/a</td>
<td>n/a</td>
<td>16 (21.6)</td>
</tr>
</tbody>
</table>
Health Data NY scores highest on indices of intrinsic data quality, contextual data quality, and adherence to Dublin Core metadata standards.
Gaps in meeting criteria from the five-star open data deployment scheme

35% of offerings meet all five criteria
Platform usability: common features

- Hosting data on platforms, with links to external pages where relevant \((Health\ Data\ NY,\ NYC\ Open\ Data)\)
- Open data handbooks to guide standardization of metadata and vocabulary \((Health\ Data\ NY,\ NYC\ Open\ Data)\)
- Multiple functions to search for and download data offerings, post comments and ideas, develop APIs, and announce innovation challenges to engage developers and the public
- Help functions such as tutorials, help email address
- Designed to engage the public, with pictures, storyboards, social media, ways for users to provide comments
- Ability to embed visualizations into external pages \((Health\ Data\ NY,\ NYC\ Open\ Data)\)
Platform usability: areas for improvement

- Healthdata.gov primarily serves as a search engine
  - All offerings hosted on external webpages, such as CDC
  - Limited interaction with data on the platform
  - Difficult to locate offerings when redirected to other sites
- Technical problems limit functionality
  - Frequent broken links *(Healthdata.gov)*
  - Problems loading map visualizations *(NYC Open Data)*
- No response to our email queries to help desks
- Low visibility on Google searches *(Healthdata.gov, NYC Open Data)*
Limitations

- New York platforms are not nationally representative
- Limited to fact-based questions (e.g. “is there a clearly identified limitations section?”)
  - Subjective nature of data quality, which depends on intended use
  - Time constraints
  - Unanticipated finding that most data objects are not tabular datasets
  - (Somewhat anticipated) finding that the three platforms present information in inconsistent formats and locations
- Coding guide does not capture:
  - Representational consistency (one aspect of platform usability)
  - Metadata consistency (one aspect of metadata quality)
- Indices need further validation
Implications for policy and practice

- Government agencies have little guidance on how to release open data for different user communities

- All three platforms have areas needing improvement, but Health Data NY scored highest by our measures

- Sustained effort on improving the usability and quality of open data is necessary for improving their value for public health

- Future work is needed to develop standard measures of quality and usability
  - Additional research on the factors that make some open data sites more successful
  - Development of checklists of “best practices” for open data managers
Other PHSSR project activities

- Key informant interviews with public health practitioners to understand the value propositions of integrating researchers into the open data ecosystem, and barriers to releasing data

- Pilot geospatial analysis of the relationship between childhood obesity and the built environment in NYS, using open data resources
  - Collaboration with Health Data NY team and Socrata
  - Comparison of results from “gold standard data ecosystem” data analysis model to: 1) no interaction with practitioners, and 2) automated platform-based findings
Questions?

Email:
emartin@albany.edu

For additional information on the PHSSR project:
www.publichealthsystems.org/erika-martin-phd-mph-0

For materials from fall 2013 workshop on open health data in New York and links to open data resources:
www.rockinst.org/ohdoo
Commentary

Guthrie Birkhead, M.D., M.P.H.
Deputy Commissioner, Office of Public Health, New York State Department of Health
Professor of Epidemiology, School of Public Health, SUNY
guthrie.birkhead@health.ny.gov

Cheryl Wold, MPH
Wold and Associates, Pasadena, California
cheryl@cherylwold.com

Questions and Discussion
Open data case studies
(practitioner commentary: Cheryl Wold)

http://stage.chcf.org/programs/marketmonitor/open-data
cheryl@cherylwold.com
Characteristics of Data Use

Data Characteristics
- Populations represented
- Sample size and sampling methods
- Unit of analysis
- Data elements included
- Data collection method
- Study design
- Data collection timing and frequency
- Data format and layout
- Amount and type of missing data
- Procedures to annotate dataset

Data User Characteristics
- Subject matter expertise
- Technical skills
- Types of tasks performed
- Intended use

Platform Promotion and User Training
- Policies, regulations, and data stewardship
- Legal interpretation of confidentiality protections
- Political support for developing and releasing data
- Capacity to respond to user feedback
- Financial resources
- Value propositions for releasing data
- Availability of information technology
- Platform advertising, promotion, and user training

Data Quality and Usability

Intrinsic Data Quality
- Accuracy+
- Believability/Reputation+
- Objectivity/Reliability+
- Confidentiality+
- Validity

Contextual Data Quality
- Relevancy+
- Value-added*
- Timeliness+
- Completeness*
- Appropriate amount of data*
- Ease of understanding+
- Ease of manipulation*
- Concise representation

Platform Usability
- Accessibility*
- Representational consistency*
- Functionality*
- User-friendliness*
- Learnability*
- Visibility*

Metadata Quality
- Completeness*
- Interpretability^+
- Accuracy^+\n- Provenance+
- Consistency*
- Timeliness
- Conformance to expectations

Legend
* Coding instrument contains at least one item to directly assess
+ Coding instrument contains at least one item to indirectly assess (e.g., “is there a clearly identified limitations section?” as an component of intrinsic data)
^ Assessed using narrative comments

Health Impacts

Short-Term Impacts
- Research studies completed
- Research grants obtained
- Development of mobile health applications
- Data-driven population health planning and monitoring
- Availability of health information
- Empowerment of healthcare consumers

Long-Term Impacts
- Quality of medical and public health services
- Value of medical and public health services
- Health status of patients and populations
- Improved decisionmaking by patients, providers, and policymakers
### Upcoming Webinars – March/April 2015

**Thursday, March 19 (1-2pm ET)**

**Cross-sector Collaboration Between Local Public Health & Health Care for Obesity Prevention**

Eduardo J. Simoes, MD, University of Missouri and
Katherine A. Stamatakis, PhD, MPH, St. Louis University

**Wednesday, April 1 (12-1pm ET)**

**Restructuring a State Nutrition Education and Obesity Prevention Program: Implications of a Local Health Department Model**

Helen W. Wu, PhD, U. California Davis — 2013 PHSSR MRDA Award

**Wednesday, April 8 (12-1pm ET)**

**Public Health Services Cost Studies: Tobacco Prevention and Mandated Public Health Services**

Pauline Thomas, MD, New Jersey Medical School & NJ Public Health PBRN
Nancy Winterbauer, PhD, East Carolina University & NC Public Health PBRN

**Tuesday and Wednesday, April 21-22**

2015 PHSSR KEENELAND CONFERENCE, Lexington, KY

---

Archives of all Webinars available at:

http://www.publichealthsystems.org/phssr-research-progress-webinars
<table>
<thead>
<tr>
<th>Date</th>
<th>Time</th>
<th>Webinar Title</th>
<th>Speaker Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wednesday, May 6</td>
<td>12-1pm ET</td>
<td>CHIP AND CHNA: MOVING TOWARDS COLLABORATIVE ASSESSMENT AND COMMUNITY HEALTH ACTION</td>
<td>Scott Frank, MD, Director, Ohio Research Association for Public Health Improvement</td>
</tr>
<tr>
<td>Wednesday, May 13</td>
<td>12-1pm ET</td>
<td>VIOLENCE AND INJURY PREVENTION: VARIATION IN PUBLIC HEALTH PROGRAM RESOURCES AND OUTCOMES</td>
<td>Laura Hitchcock, JD, Project Manager, Public Health – Seattle &amp; King County</td>
</tr>
<tr>
<td>Thursday, May 21</td>
<td>1-2pm ET</td>
<td>TDB</td>
<td></td>
</tr>
<tr>
<td>Wednesday, June 3</td>
<td>12-1pm ET</td>
<td>OPTIMIZING EXPENDITURES ACROSS THE HIV CARE CONTINUUM: BRIDGING PUBLIC HEALTH &amp; HEALTH CARE SYSTEMS</td>
<td>Gregg Gonsalves, Yale University (PPS-PHD)</td>
</tr>
<tr>
<td>Wednesday, June 10</td>
<td>12-1pm ET</td>
<td>EXAMINING PUBLIC HEALTH SYSTEM ROLES IN MENTAL HEALTH SERVICE DELIVERY</td>
<td>Jonathan Purtle, DrPH, MPH, MSc, Drexel University School of Public Health (PPS-PHD)</td>
</tr>
<tr>
<td>Thursday, June 18</td>
<td>1-2pm ET</td>
<td>INJURY PREVENTION PARTNERSHIPS TO REDUCE INFANT MORTALITY AMONG VULNERABLE POPULATIONS</td>
<td>Sharla Smith, MPH, PhD, University of Kansas School of Medicine - Wichita (PPS-PHD)</td>
</tr>
<tr>
<td>Wednesday, July 1</td>
<td>12-1pm ET</td>
<td>THE AFFORDABLE CARE ACT AND CHILDHOOD IMMUNIZATION DELIVERY IN RURAL COMMUNITIES</td>
<td>Van Do-Reynoso, University of California - Merced (PPS-PHD)</td>
</tr>
</tbody>
</table>
Thank you for participating in today’s webinar!

For more information contact:
Ann V. Kelly, Project Manager
Ann.Kelly@uky.edu
111 Washington Avenue #212
Lexington, KY 40536
859.218.2317
www.publichealthsystems.org